# Definition

## Project Overview

A key component to any strategic marketing, branding or business growth is market segmentation. The data set analyzed in this Capstone project is a user data collection from a mobile marketplace app for used goods. By segmenting the sellers into multiple groups, the company could provide a better customer support by training support staffs accordingly for each seller group when sellers reach out for assist throughout the process of selling their items.

## Problem Statement

The goal of the project is to segment the sellers into number of groups and investigate the characteristics and uniqueness of each group. The preliminary assessment by the data provider suggests that there are four distinct seller groups: top sellers, business sellers, casual sellers and new sellers.

This capstone project will further investigate the data using various techniques of clustering analysis and will determine the number of unique seller groups based on the given features of the data set.

## Metrics

The data set will be clustered after the PCA is applied and then silhouette analysis is to be conducted to visualize the clusters and to determine the optimum number of clusters.

# Analysis

## Data Exploration

Explain about data value of zero

The data was fabricated by the provider to only show the aggregated form of sellers’ activities

Histogram for price or item?

Explain abnormalities in data

* If a dataset is present for this problem, have you thoroughly discussed certain features about the dataset? Has a data sample been provided to the reader?
* If a dataset is present for this problem, are statistics about the dataset calculated and reported? Have any relevant results from this calculation been discussed?
* If a dataset is not present for this problem, has discussion been made about the input space or input data for your problem?
* Are there any abnormalities or characteristics about the input space or dataset that need to be addressed? (categorical variables, missing values, outliers, etc.)

## Exploratory Visualization

* Have you visualized a relevant characteristic or features about the dataset or input data?
* Is the visualization thoroughly analyzed and discussed?
* If a plot is provided, are the axes, title, and datum clearly defined?

## Algorithms and Techniques

PCA, Feature scaling

## Benchmark

# Methodology

## Data Preprocessing

## Implementation

## Refinement

# Results

## Model Evaluation and Validation

Split data 80/20 (at beginning) and train model with 80% and test model with the rest

## Justification

Use F-score?

# Conclusion

## Free-Form Visualization

## Reflection

## Improvement